Real-time NMR investigations of triple-helix folding and collagen folding diseases

Jean Baum[1] and Barbara Brodsky[2]

Folding of the collagen triple helix provides an opportunity to look at multichain molecular assembly. This triple helix also offers unique advantages for the study of folding because the process is very slow compared to globular proteins, and the kinetics of folding can be obtained in real time by NMR. Studies on triple-helical peptides illustrate the ability to observe kinetic folding intermediates directly and the ability to propose detailed mechanisms of folding through the use of real-time NMR methods. Defective collagen folding has been implicated in various connective tissue diseases and the capacity of NMR to look at the folding of specific sites provides a tool for obtaining information about altered folding mechanisms. Comparison of folding in peptides that model normal and diseased collagens could shed light on the molecular perturbation and the etiology of disease.

Addresses: [1]Department of Chemistry, Rutgers University, Piscataway, New Jersey 08855-0939, USA. [2]Department of Biochemistry, UMDNJ-Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, USA.

Correspondence: Jean Baum
E-mail: baum@rutchem.rutgers.edu

Electronic identifier: 1359-0278-002-R0053

Folding & Design 01 Aug 1997, 2:R53–R60

Introduction

NMR spectroscopy has become an important tool in the elucidation of protein folding intermediates and pathways [1–4]. Real-time NMR folding studies have been used to follow unfolding of proteins [5], to probe slow folding events such as proline isomerization [6–9], and to monitor transient folding intermediates that resemble a molten globule state [10]. Most NMR studies have focused on globular proteins which are composed typically of α-helices, β-sheet, and β-bends. Recently, there has been increased appreciation of the importance and widespread nature of multichain supercoiled motifs found in fibrous proteins, namely the collagen triple helix and the coiled-coil α-helix [11]. The triple helix has been recognized as a binding and structural motif in a large family of collagens and a variety of host-defense proteins [12,13]. These supercoiled systems present unusual challenges for NMR studies as a result of their multichain nature, repetitive sequence patterns, symmetry, and elongated rod-like shapes.

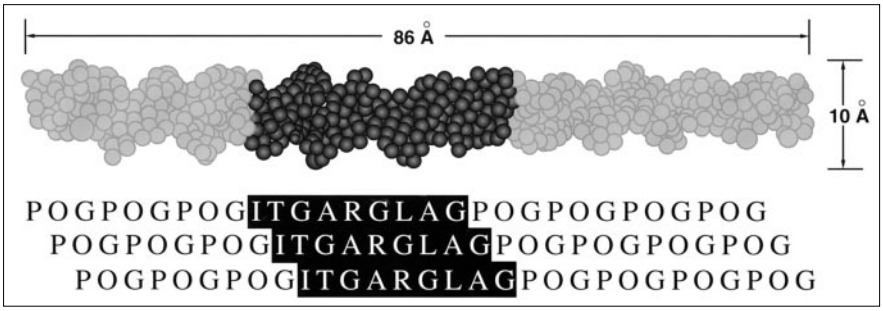
The coiled coil and triple helix provide simple systems for studying the interactions that govern folding in oligomeric proteins. Folding and association are concerted processes for coiled-coil and triple-helix fibrous proteins. Coiled-coil structures undergo this process on a millisecond to second timescale, similar to that seen for monomeric globular proteins [14,15]. In contrast, the collagen triple helix folds slowly, on a timescale of minutes to hours. The slow folding makes the triple-helix system amenable to NMR monitoring of real-time kinetics [16]. NMR studies on the collagen triple helix have been reviewed recently [17], and here we discuss the application of multidimensional NMR spectroscopy to folding in triple-helical peptides.

The collagen triple helix

The collagen triple helix is a supercoiled assembly of three extended polyproline II like chains, stabilized by a high content of the imino acids proline and hydroxyproline (Hyp, or O in the one-letter amino acid code), the presence of glycine as every third residue, and an extensive hydration network [18,19]. This leads to the repetitive sequence pattern $(\text{Gly-X-Y})_n$, where X and Y can be any amino acids but are frequently Pro and Hyp. Characterization of the sequence-dependent properties of the triple helix is critical to understanding its biological role and the interactions that determine folding. The Gly, Pro and Hyp residues are required for stabilization of the triple-helical molecule. The charged, polar, and hydrophobic residues in X and Y positions are instrumental in the formation of higher-order structures, such as fibrils and binding of other molecules. For instance, electrostatic and hydrophobic interactions are known to be critical for the assembly of collagen molecules into D-periodic fibrils [20] and similar interactions must be involved in the binding of collagen to molecules such as integrins, fibronectin, matrix metalloproteinases, and phospholipid vesicles [21]. The requirement for Gly-X-Y sequence variations to specify interactions must be balanced with sufficient imino acid content to ensure collagen stability at body temperature.

The folding of the collagen triple helix requires a transition from denatured monomers to a supercoiled helical trimer. This process has been shown to require an initial nucleation step followed by propagation [22]. The collagen molecule is synthesized in a precursor form, procollagen, which has N- and C-terminal globular propeptides. During *in vivo* folding, the globular C-propeptides facilitate chain association, registration and nucleation [23]. Nucleation of the triple helix takes place at the Gly-Pro-Hyp-rich C terminus. This is followed by a C→N-terminal zipper-like

Figure 1



Schematic representation of the collagen-like triple-helical peptide T3-785, which includes Pro-Hyp-Gly (POG) sequences at both ends and a central nine-residue sequence from human type III collagen (dark section). The dimensions of the peptide are shown, indicating the rod-like shape of the molecule. The sequence for peptide T3-785 and the one-residue stagger between the three chains are shown beneath.

propagation, which is limited by the slow *cis*/*trans* isomerization of X-Pro imide bonds [24].

Defective folding of mutant collagens has been implicated in a number of connective tissue diseases [25]. In most cases of brittle bone disease (osteogenesis imperfecta [OI]), for example, a single Gly in the $(Gly\text{-}X\text{-}Y)_n$ repeating sequence is replaced by another amino acid, and these substitutions have been shown to slow down triple-helix folding [26]. NMR studies on peptides that model these regions with mutations provide an approach for defining the folding defect at a molecular level and shedding light on the etiology of human collagen diseases.

Peptide design

Peptides serve as useful models for the collagen triple helix and are suitable for NMR studies. Using solid state peptide synthesis technology, triple-helical peptides of defined length and sequence can be made. A stable triple-helical conformation will be adopted if peptides have Gly as every third residue and a high content of imino acids. The most stabilizing Gly-X-Y tripeptide sequence is Gly-Pro-Hyp, and inclusion of $(Gly\text{-}Pro\text{-}Hyp)_n$ sequences at one or both ends of peptides has served as the basis of many peptide designs. For instance, $(Pro\text{-}Hyp\text{-}Gly)_{10}$ forms a triple helix with a melting point near 60°C in aqueous solution, while $(Pro\text{-}Hyp\text{-}Gly)_4$-Glu-Lys-Gly-$(Pro\text{-}Hyp\text{-}Gly)_5$ has a T_m near 42°C [27] and $(Pro\text{-}Hyp\text{-}Gly)_3$-Ile-Thr-Gly-Ala-Arg-Gly-Leu-Ala-Gly-$(Pro\text{-}Hyp\text{-}Gly)_4$ has a T_m near 25°C [28].

An additional strategy for stabilization has been covalent cross-linking of the three chains in the triple helix. The synthesis of peptides on a Lys-Lys dimer base has been one approach to synthesizing a variety of cross-linked trimer peptides containing simple sequences as well as biologically important collagen binding sites [29]. More recently, Goodman and colleagues have used a rigid Kemp triacid (KTA) template for attachment of three $(Gly\text{-}Pro\text{-}Hyp)_n$ chains, where the length n is varied [30,31]. The KTA approach was also applied to the incorporation of $(Gly\text{-}Pro\text{-}Nleu)_n$ chains (where Nleu is the peptoid residue N-isobutylglycine) [32]. The attachment of the peptide chains to a KTA template was shown to have a very strong stabilizing effect on the triple-helical conformation and to greatly accelerate folding.

All of the peptides described above form triple helices with three identical chains. A very significant advance has been made with the report of a heterotrimeric cross-linked triple-helical peptide that acts as a substrate for human neutrophil collagenase [33]. The design strategy utilized inclusion of cysteine residues at positions that would facilitate formation of two interstrand disulfide bridges keeping the three chains in correct register.

Triple-helical peptides that are not cross-linked, and those cross-linked on a KTA template, have proved to be highly soluble in aqueous solution and have been used for NMR conformation and dynamic studies. The folding of the triple helix is very slow for peptides that are not cross-linked and real-time NMR studies of this process have been reported on such designed peptides [16]. Cross-linking has been shown to greatly accelerate the folding of the triple helix, so that it is complete within fractions of a minute and no longer suitable for NMR investigation.

NMR studies on triple-helical peptides with repeating tripeptide sequences

Collagen-like peptides are well suited for NMR studies because of the high solubility of both the monomer and the trimer forms, their relatively low molecular weight, and their slow folding. However, the trimer form of these peptides has intrinsic features that complicate structure determination. Such features include its rod-like shape, repetitive sequence, multichain composition, and symmetry. The triple helix is an extended uniform conformation, giving rise to a rod-like molecule, so that even a 30-residue peptide has an axial ratio of greater than 8:1, as seen in Figure 1. This leads to anisotropic tumbling which results in the broadening of NMR peaks expected for a much higher molecular weight molecule and the dependence of NOE intensities on the internuclear ^1H–^1H vector and the long axis of the molecule [34,35].

Multidimensional NMR studies have been reported on simple repeating polytripeptides, such as (Pro-Hyp-Gly)$_{10}$, and shorter cross-linked peptides containing repeating Gly-Pro-Hyp or Gly-Pro-Nleu sequences [28,36–38]. Proton NMR including TOCSY, COSY and NOESY experiments have been used to obtain sequence-specific assignments for the repeating triplet unit in these peptides, but overlapping resonances prevent assignments to individual chains [28,37,38]. Because of the inability to distinguish different chains and the identical repetitive sequence within a given chain, an NOE cross-peak may arise from interactions between atoms in the same chain and/or between atoms from different chains. For example, an NOE between Gly-NH and Hyp-C_αH can arise from either an interchain or an intrachain interaction. Direct structural determination based on NOE-derived distance constraints is therefore not possible. An initial structural approach is taken by using the X-ray-derived triple-helix model to determine which NOEs arise uniquely from interchain interactions [28,37,38]. The observation of these NOEs confirms that the basic triple-helix model is present in solution in (Pro-Hyp-Gly)$_{10}$ and in the cross-linked peptides. The repetitive sequence and the stabilizing nature of Gly-Pro-Hyp triplets makes the peptide (Pro-Hyp-Gly)$_{10}$ a useful standard to compare with peptides with more varied sequences that model collagen.

NMR conformation and dynamics studies on a collagen model peptide

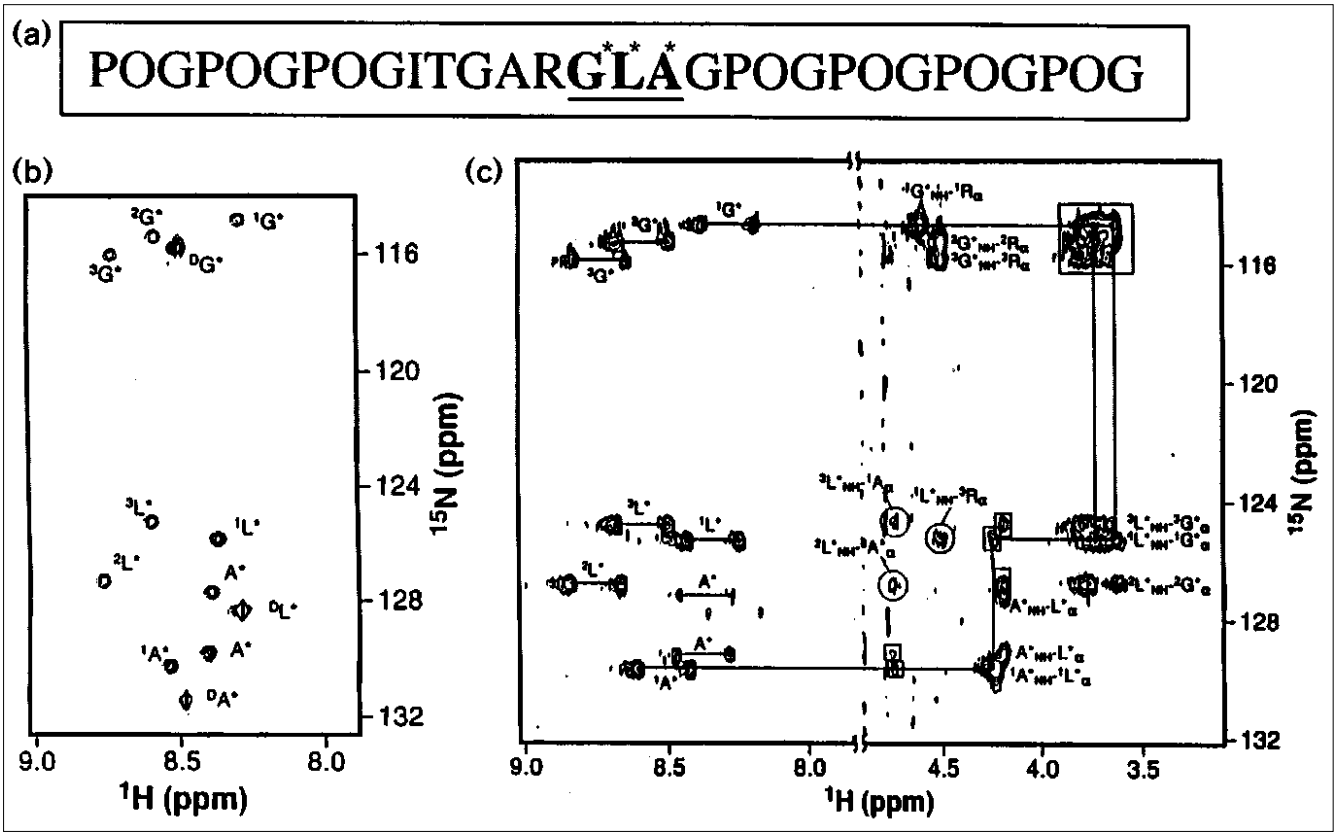
Approaches to gain more specific information about collagen through the use of NMR include the incorporation of collagen-like non-identical tripeptide sequences and the labeling of residues at specific positions in peptides (Figure 1). A collagen sequence was incorporated into peptide T3-785, the conformation, dynamics and folding of which were extensively studied by NMR [16,28,35]. Uniform labeling with ^{13}C or ^{15}N is not economically feasible on peptides and, furthermore, collagen molecules cannot be produced in bacterial expression systems because bacteria do not have the enzymes for posttranslational modification of proline to hydroxyproline. However, synthesis of peptides with ^{15}N-enriched residues at specific sites has provided an alternative approach.

Studies on T3-785 illustrate the effect of amino acid sequence on triple-helix properties and show the advantages of incorporating specifically labeled residues. Peptide T3-785 was designed to incorporate residues of type III collagen in the middle of a sequence that has Pro-Hyp-Gly caps at both ends for stabilization (Figure 1). This sequence is imino acid deficient and represents a biologically important region of collagen [28]. It is located C-terminal to the unique collagenase site (781–782) and includes the single trypsin cleavage site at residue 790. This peptide adopts a triple-helical conformation at low temperature, but is considerably less stable than (Pro-Hyp-Gly)$_{10}$.

To determine the conformation of the triple-helical peptide, one version of T3-785 was made with an ^{15}N-enriched tripeptide unit Gly-Leu-Ala [28]. The proton and ^{15}N resonances of T3-785 are assigned using HMQC, HMQC-NOESY and HMQC-TOCSY experiments [28]. The signals for the Gly, Leu and Ala residues are well resolved in the HMQC spectrum (Figure 2). Each residue shows four peaks which arise from the monomer/trimer equilibrium. One peak in the spectrum does not disappear as a function of temperature and corresponds to a monomer species. As the peptide is heated above the melting temperature, three of the peaks disappear, indicating that these derive from the triple-helix conformation. Sequential connectivities in the NOESY spectrum at low temperature can be traced for three distinct chains of the trimer, and NOEs between these chains indicate close packing in a conformation similar to that seen in the triple-helix model deduced by X-ray diffraction. This indicates that the three resonances for each residue represent three distinguishable chains in one trimer of peptide T3-785. The one-residue stagger between the three chains in the molecule results in non-equivalent environments when the triplets along the chain are nonrepetitive, as is the case for Gly-Leu-Ala in T3-785.

To further characterize the amino acid sequence dependence of the trimer state, the backbone dynamics of T3-785 and (Pro-Hyp-Gly)$_{10}$ were compared [35] (Table 1). ^{15}N relaxation and hydrogen exchange experiments were obtained for the labeled Gly-Leu-Ala residues in T3-785 and for a labeled Gly residue in the middle of the peptide (Pro-Hyp-Gly)$_{10}$. An anisotropic treatment of the data is required to obtain order parameters from the ^{15}N relaxation experiments because of the rod-like dimensions of these molecules. The values of the order parameters for the Gly in (Pro-Hyp-Gly)$_{10}$ and for the labeled residues in T3-785 are similar to one another and resemble those found in secondary structures of globular proteins. This suggests a rigidity on the picosecond timescale that is independent of sequence and stability. In contrast, hydrogen exchange rates were seen to be much faster for the Gly in T3-785 than for the Gly in (Pro-Hyp-Gly)$_{10}$. Studies on homologous sets of peptides suggest that hydrogen exchange of amide groups in the triple helix proceeds by an EX2 mechanism and reflects local stability and dynamics, rather than global stability or folding (D Siegel, J Baum, unpublished data). Variations in local dynamic properties such as seen for T3-785 and (Pro-Hyp-Gly)$_{10}$ could play a role in recognition and folding. Thus the backbone conformation shows little sequence dependence, while the dynamics are affected by sequence.

Figure 2



(a) Peptide T3-785 with an ^{15}N-labeled Gly15-Leu16-Ala17 tripeptide (shown in bold and underlined). (b) HMQC spectrum of T3-785 dissolved in 90% H_2O/10% D_2O solution (pH 2.6) at 10°C. The four peaks seen for ^{15}N-labeled Gly*, Leu*, and Ala* residues are indicated. The superscript D refers to the 'denatured' peak for each residue, and the other three peaks represent the three chains in the trimer form. For example, ^{1}G, ^{2}G, and ^{3}G are Gly from chains 1, 2, and 3, respectively. Two trimer Ala peaks were not assigned due to chemical shift overlap. (c) HMQC-NOESY spectrum of T3-785 dissolved in D_2O at 10°C, showing the NH and $C_\alpha H$ region of the spectrum. ^{15}N was not decoupled during the acquisition time. The sequential connectivities of chain 1 are fully traced out, and the sequence ^{1}R-^{1}G*-^{1}L*-^{1}A* is connected by a solid line. The boxed regions are the positions of the HMQC-TOCSY peaks. The circled peaks are the interchain NOEs.

Real-time NMR folding of a collagen model peptide

The folding of peptide T3-785 takes place over a period of hours. This slow rate of folding, together with the NMR assignments of both monomer and trimer peaks, sets the stage for direct monitoring of the kinetics of folding [16]. The folding of peptide T3-785 illustrates some of the features of real-time NMR folding of triple-helical peptides and the sequence dependence of triple-helix formation.

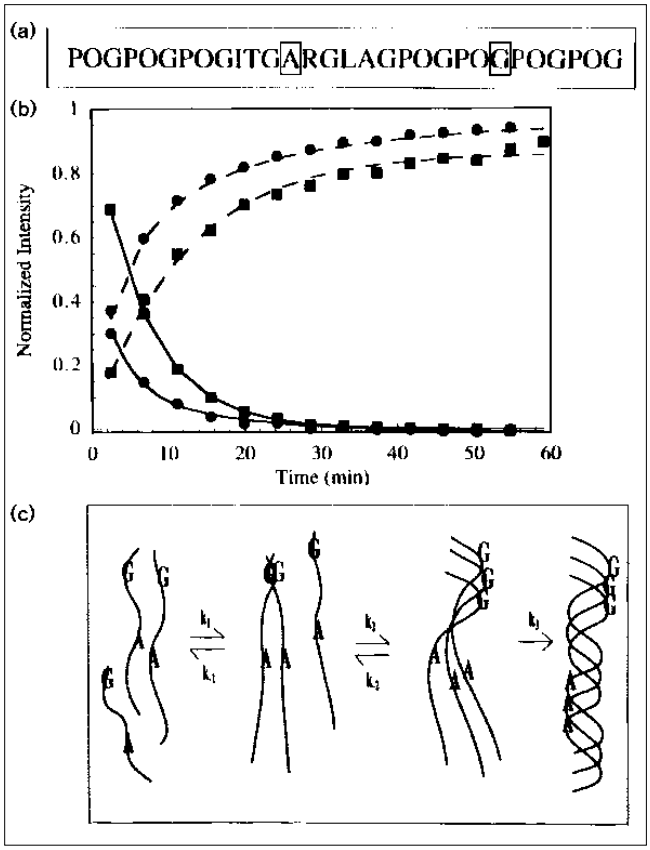
Table 1

Order parameter (S^2), hydrogen exchange rate (k_{ex}) and protection factor (P) for ^{15}N-Gly15 in the peptides T3-785 and (Pro-Hyp-Gly)$_{10}$ at 10°C [35].

Peptide	S^2	k_{ex} ($\times 10^{-4}$ min^{-1})	P
T3-785	0.85 ± 0.03	0.5	1077
(Pro-Hyp-Gly)$_{10}$	0.88 ± 0.04	< 0.04	> 10^4

Multiple peptides with labeled residues in different positions are synthesized in order to probe the directional nature of the folding process by real-time NMR. An example of one peptide with labeled residues in the central region and at the C-terminal end is shown in Figure 3a [16]. During the folding process, the kinetics of folding can be monitored by following the decrease in the intensity of the monomer peaks and/or the increase in the intensity of the trimer peaks. Figure 3b shows the kinetics of folding for spectra acquired 30 s after initiation of folding. The data show that the formation of the trimer and the loss of the monomer intensity for the Ala in the center of the chain (Ala13) is slower than for the Gly residue near the C terminus (Gly24). This disparity in initial amplitudes indicates that the C-terminal end is folding into the trimer form more rapidly than the central region, and that there is a point in the folding process where the local conformation of Gly24 is in a triple-helical form while the central Ala is still in the unfolded state.

Figure 3



(a) Peptide T3-785 with ^{15}N-labeled Gly24 and Ala13. (b) NMR folding profiles for Gly24 (filled circles) and Ala13 (filled squares). The disappearance of the monomer peaks (solid line) and the appearance of the trimer peaks (dashed line) is faster for Gly24 than for Ala13. The lines for Gly24 are second-order fits to the data and the lines for Ala13 are first-order fits to the data. (c) Proposed three-step folding mechanism for T3-785 (see text).

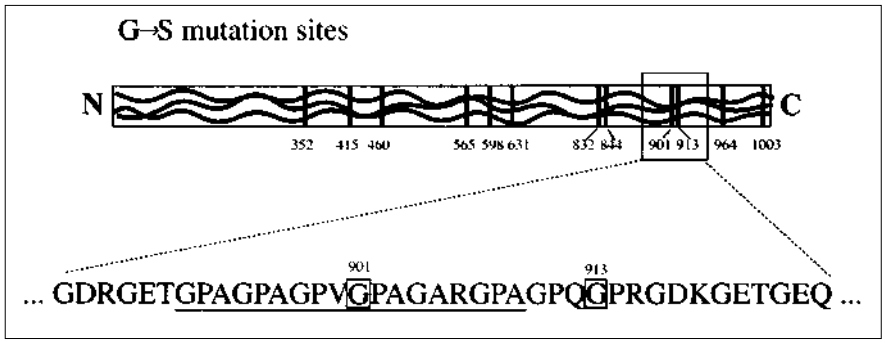
This state must represent a kinetic intermediate in the folding of triple-helical peptides. In addition, the kinetic orders in the terminal and central regions are different from one another. The C-terminal Gly folds with second-order kinetics, while the central Ala residue follows first-order kinetics.

Taken together, the NMR data are consistent with a simple mechanism of association/nucleation which is represented by the second-order kinetics of the terminal end, followed by propagation which is represented by the first-order kinetics of the central region of the peptide (Figure 3c). In the first step, the monomer chains associate to form dimers. The 'nucleated trimer' species, in which Gly24 is in a helical environment while Ala13 is still in an unfolded environment, are formed in the second step. The last step of the folding mechanism is a slow first-order propagation step where the nucleated species propagates into a triple helix. In the schematic drawing, the trimer is nucleated at one end; however, additional studies with Gly labeled at the N-terminal end indicate that nucleation can occur from either the C-terminal or N-terminal end.

The real-time NMR data on residues at different sites provide kinetic rate constants for both the nucleation and propagation steps in the folding process. These rate constants indicate that the nucleation and propagation steps are both slow in T3-785. The sequence-dependent nature of the individual steps in triple-helix folding can be investigated by NMR. Comparison of different peptides with labeled residues at comparable positions show the effect of sequence environment on folding. For example, the central Gly residue in $(Pro\text{-}Hyp\text{-}Gly)_{10}$ folds at a much faster rate compared with Gly in the same position in peptide T3-785 (J Baum, X Liu, unpublished data). One factor may be the effect of amino acid sequence on the conformation of the unfolded monomer form. The high content of sequences such as Gly-Pro-Hyp, which have a strong propensity for extended forms or β-bends, makes the denatured single chain form of triple-helical peptides unlikely to be approximated by a random coil type structure and therefore amenable to characterization by NMR [39,40]. It has been suggested that an extended conformation is more favorable for folding [41]. Preliminary NMR

Figure 4

The triple helix of type I collagen is represented at the top. Several sites of osteogenesis imperfecta (OI) Gly→Ser mutations are indicated. The region around mutations at 901 and 913 is expanded showing the amino acid sequence 892–921 of the α1 chain of type I collagen. The amino acid sequence incorporated into peptide T1-892 is underlined.



studies suggest that the monomer is nonrandom, with the central imino acid poor region in monomer T3-785 less extended than the central Pro-Hyp-Gly sequence in (Pro-Hyp-Gly)$_{10}$.

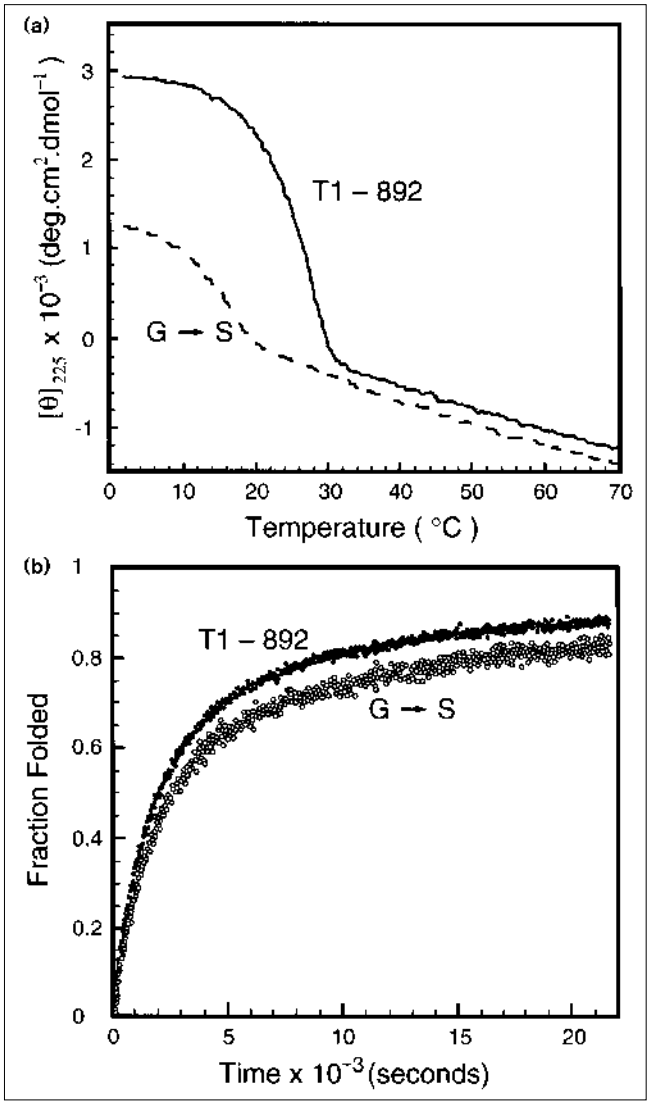
Peptide models of collagen folding diseases

NMR investigations of conformation, dynamics and folding at specific sites along the triple helix can be applied to peptides that model collagen mutations in disease. The best characterized collagen disease is osteogenesis imperfecta (OI), which results from mutations in the α1 or α2 chains of type I collagen, the major collagen in bone [25]. The majority of OI cases have a Gly→X substitution at one site as seen in Figure 4. It appears that a pathological condition results from a Gly replacement at any site along the triple-helix domain where it breaks the (Gly-X-Y)$_n$ pattern. In general, lethal mutations of OI appear to be clustered in the C-terminal 60% of the molecule, while those near the N terminus are nonlethal. This suggests a relationship to the directional propagation of the triple helix, in a C→N-terminal direction. Experimental evidence supports the concept that the Gly-substituted collagens represent folding mutations. Such evidence includes direct observation of delayed folding and increased posttranslational modification (which occurs only on unfolded chains) [25,26,42]. The nature of the folding defect remains to be defined, and NMR studies of peptides with Gly substitutions offer an opportunity to characterize folding at the substitution site, as well as other positions relative to the mutation site along the peptide.

Gly substitutions were introduced first in the highly simplified repeating model peptide (Pro-Hyp-Gly)$_{10}$ [43,44] and more recently in peptide T1-892 which contains the real sequence around an OI site [45]. Peptide T1-892 was synthesized to contain a region of the α1 chain of type I collagen where a nonlethal mutation was identified (site 901; see Figure 4). In T1-892, an 18-residue sequence of the collagen chain around the 901 mutation site was capped at the C terminus with the stabilizing (Gly-Pro-Hyp)$_4$ sequence to promote nucleation. A homologous peptide was synthesized that contained the Gly901→Ser substitution to model the case of a mild OI mutation: peptide T1-892(G→S). In CD studies (Figure 5), the thermal stability is decreased substantially by the Gly901→Ser substitution, while the folding rate was little affected. For both T1-892 and T1-892(G→S), the CD folding data fit second-order kinetics best, implying that the nucleation step is rate limiting and not affected by the Gly901→Ser substitution.

An understanding of the folding behavior of these OI peptides can be greatly enhanced by NMR real-time folding studies on peptides. Using specifically labeled residues, the nucleation and propagation steps can be discriminated and the effect of the Gly→Ser substitution

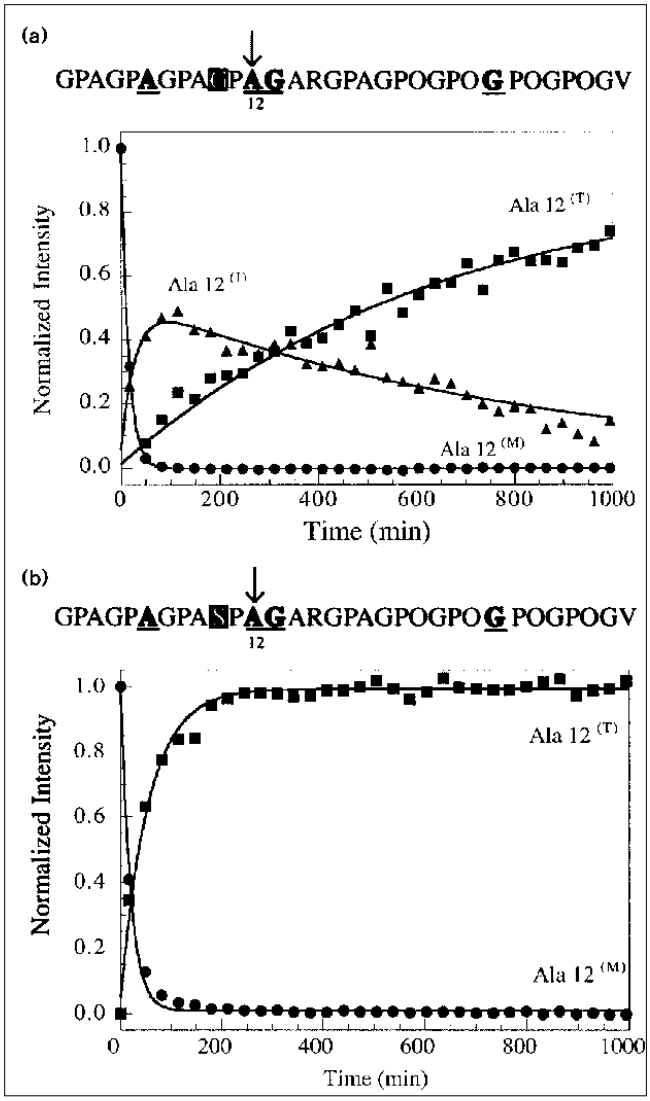
Figure 5



(a) Thermal equilibrium curves for T1-892 (solid line) and for T1-892(G→S) (dashed line) as monitored by CD at 225 nm. Both peptides contain residues 892–909 of the α1 chain of type I collagen, and one has the Gly→Ser substitution at position 901. (b) Folding kinetics of peptides T1-892 (filled circles) and T1-892(G→S) (open circles) at 2°C, pH 7. The change in ellipticity was monitored at 225 nm.

can be understood in terms of the specific steps in the folding mechanism (J Baum, B Brodsky, unpublished data) [16]. To investigate the substitution site itself, as well as residues upstream and downstream, several forms of peptides T1-892 and T1-892(G→S) were synthesized with specific ^{15}N-enriched residues (Figure 6). NMR studies indicate that for T1-892, nucleation takes place exclusively at the C terminus, followed by propagation. The unidirectional folding of T1-892 contrasts with the bidirectional folding of T3-785 and leads to an excellent

Figure 6



(a) NMR folding profile for Ala12 in peptide T1-892. The sequence of T1-892 is shown at the top; the shaded residue is the mutation site, the bold underline shows the position of ^{15}N-labeled residues in the peptide that have been monitored by NMR and the arrow indicates the ^{15}N-labeled Ala12 shown in this figure. The disappearance of the monomer peak (circles), the appearance of the trimer peak (squares) and the direct detection of a kinetic intermediate (triangles) are indicated. The lines are drawn to aid the eye. (b) NMR folding profile for Ala12 in the T1-892(G→S) peptide. The sequence is shown at the top, with the mutation site shaded, the ^{15}N-labeled positions shown as bold underline and the ^{15}N-labeled Ala12 shown in the figure indicated by an arrow. The disappearance of the monomer peak (circles) and the appearance of the trimer peak (squares) are indicated and the lines are drawn to aid the eye. In contrast to T1-892, there is no accumulation of a kinetic intermediate in the kinetic folding profile of the T1-892(G→S) peptide.

model for collagen folding. The T1-892 peptide also folds much more slowly than T3-785 and has a more complicated mechanism of folding.

In T1-892 and T1-892(G→S), NMR experiments indicate that the folding of the C-terminal residues involved in nucleation are similar to one another, but the folding of residues near the substitution site are very different from one another (J Baum, X Liu, unpublished data). In T1-892, for example, the accumulation of a kinetic intermediate at Ala12 can be detected directly in the NMR spectra and the kinetic folding profile of this intermediate along with the loss of monomer and slow trimer formation are shown in Figure 6a. From these kinetic data, it is proposed that for T1-892, at least one first-order conformational rearrangement occurs after the 'nucleated trimer' step shown in Figure 3c and before the formation of the final trimer. NMR studies on the Gly→Ser peptide indicate that it has a less ordered folded form than T1-892 (X Liu, J Baum, unpublished data), and the same Ala12 in the Gly→Ser peptide (Figure 6b) folds to its final form faster than seen for T1-892, at a rate similar to that seen for the formation of the T1-892 intermediate. Further analysis of the data should allow the definition of the folding mechanism in the mutated peptides, and the steps at which alterations are occurring.

Conclusions

Collagen triple-helical peptides offer a simple and uniform model for studying protein folding in the context of chain association and supercoiled helices. The slow folding rate of the triple helix makes it amenable to real-time NMR studies. Through the use of labeled residues at specific positions, individual steps in the folding process have been defined. The discrimination and observation of kinetic intermediates will allow clarification of the interactions that are important in chain association and propagation. Given the Gly-X-Y repeating unit of the triple helix, the identity of the X and Y residues must determine the factors important in folding. The lack of hydrophobic stabilization of the triple helix [46] makes it likely that the nature of these factors will be different than for those seen for globular proteins. In particular, it will be interesting to compare the sequence dependence of the formation of triple helices with the folding of small dimeric proteins with a more conventional hydrophobic core [47–51]. In addition to defining basic folding mechanisms, NMR studies on the triple helix have made it possible to examine perturbations at collagen mutation sites. Future studies will be directed toward clarifying the folding defects in lethal as well as nonlethal mutation sites to see whether the degree of clinical severity can be correlated with the altered folding mechanism.

Acknowledgements

We thank Don Siegel and Xiaoyan Liu for assistance in preparing the manuscript. This work was supported by NIH grants AR19626 (B Brodsky) and GM45302 (J Baum). J Baum is a Camille and Henry Dreyfus Teacher-Scholar.

References

1. Baldwin, R.L. (1993). Pulsed H/D-exchange studies of folding intermediates. *Curr. Opin. Struct. Biol.* **3**, 84–91.
2. Dobson, C.M. (1994). Solid evidence for molten globules. *Curr. Opin. Struct. Biol.* **4**, 636–640.
3. Ptitsyn, O.B. (1995). Structures of folding intermediates. *Curr. Opin. Struct. Biol.* **5**, 74–78.
4. Miranker, A.D. & Dobson, C.M. (1996). Collapse and cooperativity in protein folding. *Curr. Opin. Struct. Biol.* **6**, 31–42.
5. Kiefhaber, T., Labhardt, A.M. & Baldwin, R.L. (1995). Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* **375**, 513–515.
6. Blum, A.D., Smallcombe, S.H. & Baldwin, R.L. (1978). Nuclear magnetic resonance evidence for a structural intermediate at an early stage in the refolding of ribonuclease A. *J. Mol. Biol.* **118**, 305–315.
7. Adler, M. & Scheraga, H.A. (1988). Structural studies of a folding intermediate of bovine pancreatic ribonuclease A by continuous recycled flow. *Biochemistry* **27**, 2471–2480.
8. Akasaka, K., Naito, A. & Nakatani, H. (1991). Temperature-jump NMR study of protein folding: ribonuclease A at low pH. *J. Biomol. NMR* **1**, 65–70.
9. Koide, S., Dyson, H.J. & Wright, P.E. (1993). Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization. *Biochemistry* **32**, 12299–12310.
10. Balbach, J., Forge, V., Van Nuland, N.A.J., Winder, S.L., Hore, P.J. & Dobson, C.M. (1995). Following protein folding in real time using NMR spectroscopy. *Nat. Struct. Biol.* **2**, 865– 870.
11. Brodsky, B. (1990). Folding and higher order structure in fibrous proteins. In *The Protein Folding Problem.* (Gierasch, L. & King, J., eds.), pp. 55–62, AAAS, Washington.
12. Brodsky, B. & Shah, N.K. (1995). Folding and higher order structure in fibrous proteins. *FASEB J.* **9**, 1537–1546.
13. Cohen, C. & Parry, D.A.D. (1986). α-Helical coiled coils: a widespread motif in proteins. *Trends Biochem. Sci.* **11**, 245–248.
14. Wendt, H., Berger, C., Baici, A., Thomas, R.M. & Bosshard, H.R. (1995). Kinetics of folding of leucine zipper domains. *Biochemistry* **34**, 4097–4107.
15. Zitzewitz, J.A., Bilsel, O., Luo, J., Jones, B.E. & Matthews, C.R. (1995). Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry* **34**, 12812–12819.
16. Liu, X., Siegel, D.L., Fan, P., Brodsky, B. & Baum, J. (1996). Direct NMR measurement of the folding kinetics of a trimeric peptide. *Biochemistry* **35**, 4306–4313.
17. Mayo, K.H. (1996). NMR and X-ray studies of collagen model peptides. *Biopolymers* **40**, 359–370.
18. Bella, J., Eaton, M., Brodsky, B. & Berman, H.M. (1994). Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **266**, 75–81.
19. Rich, A. & Crick, F.H.C. (1961). The molecular structure of collagen. *J. Mol. Biol.* **3**, 483–506.
20. Hulmes, D.J.S., Miller, A., Parry, D.A.D., Piez, K.A. & Woodhead-Galloway, J. (1973). Analysis of the primary structure of collagen for the origins of molecular packing. *J. Mol. Biol.* **79**, 137–148.
21. Kadler, K. (1994). Extracellular matrix I. Fibril forming collagens. *Protein Profile* **1**, 519–638.
22. Bächinger, H.P., Bruckner, P., Timpl, R. & Engel, J. (1978). The role of *cis-trans* isomerization of peptide bonds in the coil $\Leftrightarrow$ triple helix conversion of collagen. *Eur. J. Biochem.* **90**, 605–613.
23. Prockop, D.J. & Kivirikko, K.I. (1995). Collagens: molecular biology, diseases, and potentials for therapy. *Annu. Rev. Biochem.* **64**, 403–434.
24. Engel, J. & Prockop, D.J. (1991). The zipper-like folding of collagen triple-helices and the effects of mutations that disrupt the zipper. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 137–152.
25. Byers, P.H. (1993). Osteogenesis imperfecta. In *Connective Tissue and Its Heritable Disorders: Molecular, Genetic, and Medical Aspects.* (Royce, P.M. & Steinmann, B., eds.), pp. 351–407, Wiley Liss, New York.
26. Raghunath, M., Bruckner, P. & Steinmann, B. (1994). Delayed triple helix formation of mutant collagen from patients with osteogenesis imperfecta. *J. Mol. Biol.* **236**, 940–949.
27. Venugopal, M.G., Ramshaw, J.A.M., Braswell, E., Zhu, D. & Brodsky, B. (1994). Electrostatic interactions in collagen-like triple-helical peptides. *Biochemistry* **33**, 7948–7956.
28. Li, M.-H., Fan, P., Brodsky, B. & Baum, J. (1993). Two-dimensional NMR assignments and conformation of $(Pro\text{-}Hyp\text{-}Gly)_{10}$ and a designed collagen-like triple-helical peptide. *Biochemistry* **32**, 7377–7387.
29. Fields, G.B. & Prockop, D.J. (1996). Perspectives on synthesis and applications of triple-helical collagen model peptides. *Biopolymers* **40**, 345–357.
30. Feng, Y., Melacini, G., Taulane, J.P. & Goodman, M. (1996). Acetyl-terminated and template-assembled collagen-based polypeptides composed of Gly-Pro-Hyp sequences. 2. Synthesis and conformational analysis by circular dichroism, ultraviolet absorbance, and optical rotation. *J. Am. Chem. Soc.* **118**, 10351–10358.
31. Goodman, M., Feng, Y., Melacini, G. & Taulane, J.P. (1996). A template-induced *incipient* collagen-like triple-helical structure. *J. Am. Chem. Soc.* **118**, 5156–5157.
32. Goodman, M., Melacini, G. & Feng, Y. (1996). Collagen-like triple helices incorporating peptoid residues. *J. Am. Chem. Soc.* **118**, 10928–10929.
33. Ottl, J., *et al.*, & Moroder, L. (1996). Design and synthesis of heterotrimeric collagen peptides with a built-in cystine knot. *FEBS Lett.* **398**, 31–36.
34. Withka, J.M., Swaminathan, S. & Bolton, P.H. (1990). NOEs in duplex DNA depend on orientations of internuclear vectors to the symmetry axis. *J. Magn. Res.* **89**, 386–390.
35. Fan, P., Li, M.-H., Brodsky, B. & Baum, J. (1993). Backbone dynamics of $(Pro\text{-}Hyp\text{-}Gly)_{10}$ and a designed collagen-like triple helical peptide by ^{15}N NMR relaxation and hydrogen exchange measurements. *Biochemistry* **32**, 13299–13309.
36. Brodsky, B., Li, M.-H., Long, C.G., Apigo, J. & Baum, J. (1992). NMR and CD studies of triple-helical peptides. *Biopolymers* **32**, 447–451.
37. Melacini, G., Feng, Y. & Goodman, M. (1996). Acetyl-terminated and template-assembled collagen-based polypeptides composed of Gly-Pro-Hyp sequences. 3. Conformational analysis by ^1H-NMR and molecular modeling studies. *J. Am. Chem. Soc.* **118**, 10359–10364.
38. Melacini, G., Feng, Y. & Goodman, M. (1996). Collagen-based structures containing the peptoid residue N-isobutylglycine (Nleu). 2. Conformational analysis of Gly-Pro-Nleu sequences by ^1H NMR, CD, and molecular modeling. *J. Am. Chem. Soc.* **118**, 10725–10732.
39. Dyson, H.J. & Wright, P.E. (1993). Peptide conformation and protein folding. *Curr. Opin. Struct. Biol.* **3**, 60–65.
40. Mayo, K.H., Parra-Diaz, D., McCarthy, J.B. & Chelberg, M. (1991). Cell adhesion promoting peptide GVKGDKGNPGWPGAP from collagen type IV triple helix: *cis/trans* proline-induced multiple ^1H NMR conformations and evidence for a KG/PG multiple turn repeat motif in the all-*trans* proline state. *Biochemistry* **30**, 8251–8267.
41. Chopra, R.K. & Ananthanaryanan, V.S. (1982). Conformational implications of enzymatic proline hydroxylation in collagen. *Proc. Natl Acad. Sci. USA* **79**, 7180–7184.
42. Bächinger, H.P., Morris, N.P. & Davis, J.M. (1993). Thermal stability and folding of the collagen triple helix and the effects of mutations in osteogenesis imperfecta on the triple helix of type I collagen. *Am. J. Med. Gen.* **45**, 152–162.
43. Long, C.G., Braswell, E., Zhu, D., Apigo, J., Baum, J. & Brodsky, B. (1993). Characterization of collagen-like peptides containing interruptions in the repeating Gly-X-Y sequence. *Biochemistry* **32**, 11688–11695.
44. Long, C.G., Li, M.-H., Baum, J. & Brodsky, B. (1992). Nuclear magnetic resonance and circular dichroism studies of a triple-helical peptide with a glycine substitution. *J. Mol. Biol.* **225**, 1–4.
45. Yang, W., Battinenni, M. & Brodsky, B. (1997). Amino acid sequence environment modulates the disruption by osteogenesis imperfecta glycine substitutions in collagen-like peptides. *Biochemistry* **36**, 6930–6935.
46. Shah, N.K., Ramshaw, J.A.M., Kirkpatrick, A., Shah, C. & Brodsky, B. (1996). A host-guest set of triple helical peptides: stability of Gly-X-Y triplets containing common nonpolar residues. *Biochemistry* **35**, 10262–10268.
47. Milla, M.E. & Sauer, R.T. (1994). P22 Arc repressor: folding kinetics of a single-domain, dimeric protein. *Biochemistry* **33**, 1125–1133.
48. Waldburger, C.D., Jonsson, T. & Sauer, R.T. (1996). Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl Acad. Sci. USA* **93**, 2629–2634.
49. Mok, Y.-K., Bycroft, M. & De Prat-Gay, G. (1996). The dimeric DNA binding domain of the human papillomavirus E2 protein folds through a monomeric intermediate which cannot be native-like. *Nat. Struct. Biol.* **3**, 711–717.
50. Munson, M., Anderson, K.S. & Regan, L. (1997). Speeding up protein folding: mutations that increase the rate at which Rop folds and unfolds by over four orders of magnitude. *Fold. Des.* **2**, 77–87.
51. Gittelman, M.S. & Matthews, C.R. (1990). Folding and stability of *trp* aporepressor from *Escherichia coli*. *Biochemistry* **29**, 7011–7020.